



DIMERIC FLUORESCENT POLYPEPTIDES

#6

INS
ACI

BACKGROUND OF THE INVENTION

Fluorescent proteins are widely used in the fields of biochemistry, molecular and cell biology, medical diagnostics and drug screening methodologies (Chalfie et al., 1994, Science 5 263: 802-805; Tsien, 1998, Ann. Rev. Biochem. 67: 509-544). One property shared by the most useful fluorescent proteins is that they require no host-encoded co-factors or substrates for fluorescence. The proteins therefore retain their fluorescent properties both in isolation from their native organism, and when expressed in the cells of other organisms. This property makes them particularly well suited for a variety of in vivo and in vitro applications. Another major 10 advantage of fluorescent proteins for use in biological systems is that they are indeed proteins, which permits their synthesis, within cells or organisms of interest, avoiding a host of problems relating to the attachment of the label to a protein of interest and/or delivery of labeled proteins into a cell. Not only can the proteins be made within the desired cell or organism, but they also retain their fluorescent properties when expressed as fusions with other proteins of interest, 15 which greatly enhances their utility both in vivo and in vitro.

Fluorescent proteins have been used as reporter molecules to study gene expression in culture as well as in transgenic animals by insertion of fluorescent protein coding sequences downstream of an appropriate promoter. They have also been used to study the subcellular localization of proteins by direct fusion of test proteins to fluorescent proteins, and fluorescent 20 proteins have become the reporter of choice for monitoring the infection efficiency of viral vectors both in cell culture and in animals. Variants of fluorescent proteins exhibiting spectral shifts in response to changes in the cellular environment (e.g., changes in pH, ion flux, or the

redox status of the cell) are also used to monitor such changes (see, for example, Inouye & Tsuji, 1994, FEBS Lett. 351: 211-214; Miyawaki et al., 1997, Nature 388: 882-887).

Perhaps the most promising role for fluorescent proteins as biochemical markers is their application to methods that exploit fluorescence resonance energy transfer (FRET). FRET occurs with fluorophores for which the emission spectrum of one fluorophore overlaps with the excitation spectrum of a second fluorophore. When such fluorophores are brought into close proximity, excitation of the “donor” fluorophore results in emission from the “acceptor”. Pairs of such fluorophores are thus useful for monitoring molecular interactions. Fluorescent proteins are useful for the analysis of protein:protein molecular interactions in vivo or in vitro if their respective fluorescent emission and excitation spectra overlap to allow FRET. The donor and acceptor fluorescent proteins may be produced as fusions with the proteins one wishes to analyze for interactions. These types of applications of fluorescent proteins are particularly appealing for high throughput analyses, since the readout is direct and independent of subcellular localization.

The prototypical fluorescent protein is the *Aequorea victoria* green fluorescent protein (GFP), which was the first green fluorescent protein cloned (Prasher et al., 1992, Gene 111: 229-233). Purified *A. victoria* GFP is a monomeric protein of about 27 kDa that absorbs blue light with an excitation wavelength maximum of 395 nm, with a minor peak at 470 nm, and emits green fluorescence with an emission wavelength of about 510 nm and a minor peak near 540 nm (Ward et al., 1979, Photochem. Photobiol. Rev. 4: 1-57). The polypeptide has several drawbacks, including relatively broad excitation and emission spectra, low quantum yield, and low expression in cells of higher eukaryotes. Mutants with improved spectral characteristics and higher quantum yield have been identified, and expression in higher eukaryotes has been

improved by “humanizing” the nucleic acid sequences to encode codons optimized for human or mammalian expression.

Additional fluorescent proteins include, but are not limited to those expressed by *Discosoma sp.* and *Phialidium gregarum* (Ward et al., 1982, Photochem. Photobiol. 35: 803-808; Levine et al., 1982, Comp. Biochem. Physiol. 72B:77-85). Also, *Vibrio fischeri* strain Y1 expresses a yellow fluorescent protein that requires flavins as a co-factor for its fluorescence (Baldwin et al., 1990, Biochemistry 29: 5509-5515).

Additional cloned fluorescent proteins include, for example, the green fluorescent proteins from the sea pansy, *Renilla mullerei* (WO/99/49019) and from *Renilla reniformis* (see SEQ ID NO: 1; Figure 1). Each of these fluorescent proteins and others are useful for a variety of in vivo and in vitro uses. The *R. reniformis* GFP (rGFP) clone is particularly important, since rGFP is seen as the benchmark protein among known naturally-occurring fluorescent proteins. rGFP has 3 to 6-fold higher quantum yield than *A. victoria* GFP, and the excitation and emission spectra are narrower, making rGFP more suitable for applications involving, for example, FRET.

One major drawback shared by the GFPs from *A. victoria*, *R. mullerei* and *R. reniformis*, as well as by all known variants of those proteins, is that they are dimeric. Generally, the proteins exist as homodimers. However, when more than one form of a given GFP is expressed in a single cell or is mixed in vitro, heterodimers can form if the dimerization interfaces for the different fluorescent proteins are complementary. Heterodimerization interferes with the usefulness of fluorescent proteins for several reasons.

First, heterodimerization is undesirable when fluorescent proteins are used in energy transfer-based analyses because heterodimerization raises the background of acceptor fluorescence without a real interaction between the proteins or protein domains of interest.

When FRET is used, for example to monitor protein:protein interactions, donor and acceptor fluorescent fusion proteins are often expressed in the same cell or otherwise mixed. In the absence of heterodimerization, the excitation of the donor fluorophore leads to emission by the acceptor fluorophore only if the two fusion proteins are in close apposition. However, if
5 heterodimerization occurs between the differing fluorescent proteins (e.g., between a wild-type rGFP and an rGFP variant that is a fluorescence donor to the wild-type GFP), excitation of the donor will result in emission by the acceptor regardless of the interaction between the fused polypeptides being examined for interaction. This generates an unacceptably high background fluorescence from the acceptor fluorophore.

10 Another problem caused by the heterodimerization is that the dimerization interfaces between the proteins can serve to artifactually bring fusion polypeptides linked to the fluorescent protein monomers into close contact. The inappropriate recruitment of proteins into close apposition can have biological consequences that make data interpretation difficult. For example, some cell surface receptors gain the ability to initiate an intracellular signaling cascade
15 following ligand-induced dimerization. If the dimerization interfaces of the fluorescent proteins inappropriately recruit the fused receptor monomers into close contact, the signaling cascade can be inappropriately initiated in the absence of ligand. There is a need in the art for fluorescent proteins that do not heterodimerize.

U.S. Patent No. 5,981,200 (Tsien et al.) teaches donor and acceptor fluorescent proteins
20 linked by a peptide linker. The linked donor and acceptor proteins, referred to as “tandem fluorescent proteins,” are taught to be useful for assaying enzymes capable of cleaving the linker peptide sequence. When linked, the tandem fluorescent proteins exhibit either no fluorescence (e.g., when one protein quenches the fluorescence of the other) or fluorescence characteristic of

the acceptor. Following cleavage, the fluorescence emitted is that characteristic of the individual fluorescent proteins. Assays using this arrangement will not work unless the tandem fluorescent proteins are related as donor and acceptor.

SUMMARY OF THE INVENTION

5 The invention encompasses a recombinant fusion polypeptide comprising a first polypeptide peptide bonded to a second polypeptide, wherein the first and second polypeptides are found in nature as monomers of a multimeric protein, and wherein the recombinant fusion polypeptide is fluorescent when exposed to light of an excitation wavelength or when interactive with an excited donor fluorophore.

10 In one embodiment, the first polypeptide and the second polypeptide are peptide bonded to each other via a linker sequence.

 In another embodiment, the recombinant fusion polypeptide further comprises a third polypeptide peptide bonded to the recombinant fusion polypeptide. The third polypeptide can be peptide bonded to the recombinant fusion polypeptide either directly or through a peptide linker
15 sequence. A recombinant fusion polypeptide of this embodiment is referred to in this summary as a “fluorescent polypeptide fusion.” In a preferred embodiment, the third polypeptide is fused to the amino terminus of the first polypeptide. In another preferred embodiment, the third polypeptide is fused to the carboxy terminus of the second polypeptide sequence.

 In an additional preferred embodiment, the third polypeptide is a member of a specific
20 binding pair.

 In another embodiment, one or both of the first and second polypeptides is a monomer of one of *R. reniformis* GFP, *R. mulleri* GFP or *A. victoria* GFP.

In another embodiment, both of the first and second polypeptides are a monomer of one of *R. reniformis* GFP, *R. mulleri* GFP or *A. victoria* GFP.

The invention further encompasses a polynucleotide encoding, a recombinant fusion polypeptide comprising a first polypeptide peptide bonded to a second polypeptide, wherein the first and second polypeptides are found in nature as monomers of a multimeric protein, and wherein the recombinant fusion polypeptide is fluorescent when exposed to light of an excitation wavelength or when interactive with an excited donor fluorophore.

10 In one embodiment, the first polypeptide and the second polypeptide encoded by the polynucleotide are peptide bonded to each other via a linker sequence. In a preferred embodiment, the linker sequence encoded by the polynucleotide is from 5 to 50 amino acids long. In a further preferred embodiment, the linker sequence comprises one or more iterations of a peptide, for example the peptide RARDPRVPVAT (i.e., Arg-Ala-Arg-Asp-Pro-Arg-Val-Pro-Val-Ala-Thr). In a further preferred embodiment, the linker sequence is selected from the group consisting of (Arg-Ala-Arg-Asp-Pro-Arg-Val-Pro-Val-Ala-Thr)_n, (Gly-Ser)_n, (Thr-Ser-Pro)_n,
15 (Gly-Gly-Gly)_n, and (Glu-Lys)_n, wherein n is 1 to 15.

In another embodiment, the polynucleotide further encodes a third polypeptide peptide bonded to the recombinant fusion polypeptide. The third polypeptide encoded by the polynucleotide may be joined directly or via an encoded peptide linker.

20 In a preferred embodiment, the third polypeptide encoded by the polynucleotide is a member of a specific binding pair. It alternatively preferred that the third encoded polypeptide is fused to the amino terminus of the first polypeptide. Is additionally preferred that the third encoded polypeptide is fused to the carboxy terminus of the second polypeptide.

In another preferred embodiment, one or both of the first and second polypeptides is a monomer of one of *R. reniformis* GFP, *R. mulleri* GFP, *A. victoria* GFP.

In another preferred embodiment, both of the first and second polypeptides is a monomer of one of *R. reniformis* GFP, *R. mulleri* GFP, *A. victoria* GFP.

5 The invention further encompasses a vector comprising a polynucleotide encoding a recombinant fusion polypeptide comprising a first polypeptide peptide bonded to a second polypeptide, wherein the first and second polypeptides are found in nature as monomers of a multimeric protein, and wherein the recombinant fusion polypeptide is fluorescent when exposed to light of an excitation wavelength or when interactive with an excited donor fluorophore.

10 The invention further encompasses a cell comprising a vector comprising a polynucleotide encoding a recombinant fusion polypeptide comprising a first polypeptide peptide bonded to a second polypeptide, wherein the first and second polypeptides are found in nature as monomers of a multimeric protein, and wherein the recombinant fusion polypeptide is fluorescent when exposed to light of an excitation wavelength or when interactive with an
15 excited donor fluorophore.

In one embodiment, the cell is a bacterial cell.

In another embodiment, the cell is a eukaryotic cell. In a preferred embodiment, the eukaryotic cell is a yeast cell, an insect cell, or a mammalian cell.

20 The invention further encompasses a pair of polypeptides comprising a polypeptide labeled with a fluorescent dye and a recombinant fusion polypeptide comprising a first polypeptide peptide bonded to a second polypeptide, wherein the first and second polypeptides are found in nature as monomers of a multimeric protein, wherein the fusion polypeptide is fluorescent when exposed to light of an excitation wavelength or when interactive with an

excited donor fluorophore, and wherein the fluorescent dye and the recombinant fusion polypeptide are fluorescent donor and acceptor to each other.

The invention further encompasses a pair of recombinant fusion polypeptides comprising (a) a first fusion polypeptide comprising a first polypeptide peptide bonded to a second polypeptide, wherein the first and second polypeptides are found in nature as monomers of a multimeric protein, and wherein the first fusion polypeptide is fluorescent when exposed to light of an excitation wavelength or when interactive with an excited donor fluorophore, and (b) a second fusion polypeptide comprising a third polypeptide peptide bonded to a fourth polypeptide, wherein the third and fourth polypeptides are found in nature as monomers of a multimeric protein, and wherein the second fusion polypeptide is fluorescent when exposed to light of an excitation wavelength or when interactive with an excited donor fluorophore, wherein the first fusion polypeptide and the second fusion polypeptide are fluorescent donor and acceptor to each other.

In one embodiment, each of the first and second fusion polypeptides further comprises an additional fused (third) polypeptide, wherein the additional fused polypeptide of the first fusion polypeptide comprises a sequence which is different from the additional fused polypeptide of the second fusion polypeptide.

The invention further encompasses a method of producing a fluorescently labeled recombinant fusion polypeptide, the method comprising the steps of introducing to a cell a polynucleotide encoding a recombinant fusion polypeptide comprising a first polypeptide peptide bonded to a second polypeptide, wherein the first and second polypeptides are found in nature as monomers of a multimeric protein, and wherein the recombinant fusion polypeptide is fluorescent when exposed to light of an excitation wavelength or when interactive with an

excited donor fluorophore, and culturing the cell under conditions that permit the synthesis of the recombinant fusion polypeptide, whereby the recombinant fusion polypeptide is produced.

The invention further encompasses a method of labeling a cell with a fluorescent recombinant fusion polypeptide, the method comprising the steps of: a) introducing to a cell a polynucleotide encoding a recombinant fusion polypeptide comprising a first polypeptide peptide bonded to a second polypeptide, wherein the first and second polypeptides are found in nature as monomers of a multimeric protein, and wherein the recombinant fusion polypeptide is fluorescent when exposed to light of an excitation wavelength or when interactive with an excited donor fluorophore; and b) culturing the cell under conditions that permit the synthesis of the recombinant fusion polypeptide, whereby the cell is labeled with the fluorescent recombinant fusion polypeptide.

In a preferred embodiment, in the introducing step (a), the polynucleotide introduced to the cell further comprises a sequence encoding a third polypeptide fused in frame to the sequence encoding the recombinant fusion polypeptide.

15 The invention further encompasses a method of monitoring the interaction of two polypeptides of interest, the method comprising the steps of: a) contacting a fluorescent polypeptide fusion, as described above, and a second polypeptide wherein: i) the fluorescent polypeptide fusion comprises a first polypeptide of interest; ii) the second polypeptide comprises a second polypeptide of interest and is fluorescently labeled;, and iii) the fluorophores comprised
20 by the fluorescent polypeptide fusion and the second polypeptide are fluorescent donor and fluorescent acceptor to each other; b) exciting the donor fluorophore; and c) detecting fluorescent emission from the fluorescent acceptor, wherein the emission is indicative of the interaction of the first and the second polypeptides of interest.

In one embodiment, the second polypeptide comprises a second fluorescent polypeptide fusion, as described above, wherein the polypeptide of interest of the second fluorescent polypeptide fusion is different from the polypeptide of interest of the first fluorescent polypeptide fusion.

5 In one embodiment, the contacting step is performed in vitro.

In another embodiment, the contacting step is performed in a cell. In a preferred embodiment, the contacting comprises the step of introducing nucleic acid encoding the polypeptides to a cell.

The invention further encompasses a method of screening for a compound that modulates the interaction of a first and a second member of a specific binding pair, the method comprising the steps of: a) contacting a first polypeptide and a second polypeptide in the presence and absence of a candidate modulator wherein: i) the first polypeptide is a fluorescent polypeptide fusion, as described above, wherein the third polypeptide is the first member of a specific binding pair; ii) the second polypeptide is fluorescently labeled and comprises the second member of a specific binding pair; and iii) the fluorophores comprised by the first and second polypeptides are fluorescent donor and acceptor to each other; b) exciting the donor fluorophore; and c) detecting the fluorescence of the acceptor fluorophore, wherein emission of the spectrum characteristic of the fluorescent acceptor indicates the interaction of the first and the second members of the specific binding pair, and wherein a change in the interaction in the presence of the candidate modulator indicates that the candidate modulator modulates the interaction of the members of the specific binding pair.

In one embodiment, the second polypeptide is a fluorescent polypeptide fusion, as described above, which comprises the second member of a specific binding pair.



The phrase “fluorescent when excited” means that when the recombinant fusion polypeptide is exposed to light of an excitation wavelength or when the polypeptide interacts with an excited donor fluorophore, the polypeptide fluoresces. “Exposed to light of an excitation wavelength” means irradiated with light (electromagnetic radiation) within a given spectrum of wavelengths that is absorbed by the polypeptide such that the polypeptide emits light having a different spectrum of wavelengths, and thus fluoresces. Fluorescent emission occurs at a longer wavelength than does excitation.

A recombinant fusion polypeptide according to the invention has three properties: 1) it must emit light upon irradiation with light of a given wavelength or wavelengths; 2) it must have the capacity to form an intramolecular homodimer as defined herein above; and 3) the first and second polypeptide monomers that constitute the fusion polypeptide cannot function as fluorescent donor and fluorescent acceptor, respectively, in the context of fluorescence resonance energy transfer.

As used herein, the term “light of an excitation wavelength” refers to those wavelengths of light that are absorbed by and excite a given fluorophore to emit fluorescence. These wavelengths are described in detail herein below. Light of an appropriate portion of the spectrum is synonymous with light within the excitation spectrum of a given fluorophore.

As used herein, the term “excited donor fluorophore” refers to a fluorophore which has absorbed energy within its excitation spectrum. An excited donor fluorophore can transmit energy sufficient to excite an acceptor fluorophore.

As used herein, the term “fluorescent dye” refers to a non-polypeptide chemical moiety that, upon absorption of light energy of a particular wavelength or wavelengths, emits light at

another wavelength or that emits light when paired with an appropriate excited donor fluorophore.

When referring to members of a pair of fluorophores (i.e., fluorescent dyes or polypeptides) that can undergo fluorescence resonance energy transfer (FRET), the fluorophore that emits at a wavelength or spectrum of wavelengths that excites the other member of the pair is referred to as the “fluorescent donor” or “fluorescence donor”. Conversely, the member of the pair that emits in response to excitation by the fluorescence donor is termed the “fluorescent acceptor” or “fluorescence acceptor”. The members of such a pair are said to be “fluorescent donor and acceptor to each other.” According to the invention, the fluorescence donor and fluorescence acceptor polypeptides are not linked by peptide bonds. In one embodiment of the invention, either of the fluorescence donor or acceptor, but not both, may be a non-polypeptide fluorescent dye (also not covalently linked to each other).

As used herein, the term “fluorescently labeled” means, when referring to a polypeptide, that the polypeptide is covalently attached to a fluorescent moiety. A polypeptide may be fluorescently labeled by covalent attachment to a non-polypeptide fluorescent dye, or alternatively, by expression as a fusion protein with a fluorescent polypeptide.

In nature and as used herein, a fluorescent polypeptide is distinguished from a luminescent polypeptide in that a fluorescent polypeptide requires an input of electromagnetic energy in order to emit light, while a luminescent polypeptide emits light in response to release of chemical energy. A luminescent polypeptide may serve as a donor of excitation, energy for a fluorescent polypeptide (in fact, this is exactly what happens in nature when, for example, *Renilla* luciferase emits energy that excites *Renilla* GFP). A fusion polypeptide according to the invention may or may not be luminescent.

As used herein, the term "linker sequence" refers to a sequence of peptide, bonded amino acids that joins or links by peptide bonds two amino acid sequences or polypeptide domains that are not joined by peptide bonds in nature. A linker sequence is encoded in frame on a polynucleotide between the sequences encoding the two polypeptide domains joined by the linker. A linker is preferably 5 to 50 amino acids in length, more preferably 10 to 20 amino acids in length. An example of linkers useful in the invention are the Gly-Ala linkers taught by Huston et al., U.S. Patent No. 5,258,498, incorporated herein by reference. Additional useful linkers include, but are not limited to (Arg-Ala-Arg-Asp-Pro-Arg-Val-Pro-Val-Ala-Thr)₁₋₅ (Xu et al., 1999, Proc. Natl. Acad. Sci. U.S.A. 96: 151-156), (Gly-Ser)_n (Shao et al., 2000, Bioconjug. Chem. 11:822-826), (Thr-Ser-Pro)_n (Kroon et al., 2000, Eur. J. Biochem. 267: 6740-6752), (Gly-Gly-Gly)_n (Kluczyk et al., 2000, Peptides 21: 1411-1420), and (Glu-Lys)_n (Klyczyk et al., 2000, supra), wherein n is 1 to 15.

As used herein, the term "specific binding pair" refers to a pair of polypeptides that physically interact in a specific manner that gives rise to a biological activity, that is, to the substantial exclusion of other polypeptides. Members of a specific binding pair interact through complementary interaction domains, such that they interact to the substantial exclusion of proteins that do not have a complementary interaction domain. Non-limiting examples of specific binding pairs include antibody-antigen pairs, enzyme-substrate pairs, dimeric transcription factors (e.g., AP-1, composed of Fos specifically bound to Jun via a leucine zipper interaction domain) and receptor-ligand pairs.

As used herein, the term "amino terminus" refers to the last amino acid at the amino end of a polypeptide, where the last amino acid is not peptide bonded to another amino acid.

As used herein, the term “labeling a cell” refers to the expression of a fluorescent polypeptide in a cell, such that the cell is detectable by irradiating the cell with light within the excitation spectrum of the fluorescent polypeptide and monitoring or detecting emission within the emission spectrum of the polypeptide. A cell may be labeled by expression of a fluorescent polypeptide that localizes anywhere in the cell, including, but not limited to the cell surface, the cytoplasm, the nucleus or to particular organelles such as mitochondria, lysosomes, endosomes, golgi apparatus, endoplasmic reticulum or other specific sub-cellular locale.

10 As used herein, the term “introducing a nucleic acid into a cell” or “introducing a polynucleotide into a cell” refers to the process whereby a recombinant polynucleotide is put into a cell. Methods for introducing a nucleic acid to a cell will vary with the nature of the cell and the nature of the chosen vector, but one of skill in the art may readily select and employ a known method appropriate for a given cell type and vector.

20 IPTG or tetracycline) conducive to the synthesis of the recombinant polypeptide. One of skill in the art is assumed to be capable of maintaining yeast, insect, mammalian or other cells under conditions that permit the synthesis of a recombinant polypeptide according to the invention.

As used herein, the term “monitoring the interaction” refers to the process whereby the physical association of two polypeptides or a polypeptide and another entity are measured. As relates to the invention, the term refers most frequently to detection or measurement of association or interaction using FRET.

5 As used herein, the term “intramolecular dimer” refers to a dimer formed by the covalent peptide linkage of two polypeptide monomers. An “intramolecular dimer fluorescent protein” (IDFP) is an intramolecular dimer in which the linked polypeptides which exist in nature as monomers of a multimeric protein are fluorescent polypeptides. According to the invention, the linked monomers of an IDFP are not fluorescent donor and acceptor to each other. An “IDFP
10 fusion protein” is an IDFP which is fused to a protein of interest or to a fragment of a protein of interest.

As used herein, the term “protein of interest” refers to a polypeptide, or a domain (fragment) of a polypeptide, that is selected to be fused to an IDFP. Any polypeptide or fragment of a polypeptide for which a polynucleotide sequence is known can be fused to an
15 IDFP by standard techniques known in the art. A protein of interest according to the invention either does not alter the fluorescence characteristics of the fused IDFP, or, if it does alter those characteristics, the alteration is such that the alteration does not interfere with the intended use of the IDFP fusion protein.

As used herein, the term “detecting fluorescence” refers to the process whereby the
20 fluorescent emission by a fluorescent polypeptide is measured or determined. Fluorescence detection methods include quantitative and qualitative methods adapted for standard or confocal microscopy, FACS analysis, and those adapted for high throughput methods involving multiwell plates, arrays or microarrays. One of skill in the art can select appropriate filter sets and

excitation energy sources for the detection of fluorescent emission from a given fluorescent polypeptide or dye.

As used herein, the term “candidate modulator” refers to an agent being evaluated for its effect on the function of a polypeptide or the interaction of members of a specific binding pair.

5 Exemplary sources and types of candidate modulators useful according to the invention are described herein below.

As used herein, the term “change in interaction” or “modulation of interaction” refers to an increase or decrease in the level of interaction detected between members of a specific binding pair. As used herein, the level of interaction is considered increased if the detected
10 interaction goes up by at least 10%, and preferably by 20%, 35%, 50%, 75%, or more, up to and including 2-fold, 5-fold, 10-fold, 20-fold, 50-fold or more relative to a standard. As used herein, the level of interaction is considered decreased if the detected interaction goes down by at least 10%, and preferably by 20%, 35%, 50%, 75%, 90%, 95%, 98%, 99% or more, up to and including 100% (no interaction) relative to a standard.

15 As used herein, the term “single polypeptide chain” refers to a polypeptide chain in which all amino acids are linked sequentially by peptide bonds. A “single polypeptide chain” is one generated by translation of a single mRNA template and may encompass one or more polypeptide domains, including one or more repeats of the sequence comprising one polypeptide or polypeptide domain.

20 As used herein, the term “polypeptide domain” refers to a sequence of amino acids that exhibits one or more discrete binding or functional properties. As used herein, binding or functional properties include binding to one or more polypeptides, modulation of the binding of one or more polypeptides, recognition by an antibody or antigen binding fragment thereof

preferably less than 0.01%, 0.001% or fewer monomers are physically associated with polypeptides that do not have a complementary interaction domain.

As used herein, the term “variant” refers to a polypeptide that differs in amino acid, sequence from a parent polypeptide yet retains the function of the parent polypeptide. A variant fluorescent polypeptide may, for example, have one or more amino acid insertions, deletions or substitutions that do not alter ability of the polypeptide to emit fluorescence upon excitation or interaction with a donor or acceptor fluorophore. A variant fluorescent polypeptide according to the invention has the ability to form an intramolecular homodimer as defined herein.

As used herein, the term “derived from” refers to a polypeptide that differs in amino acid sequence from a reference polypeptide used as the template or starting sequence for generating or deriving the differing sequence. For example, a fluorescent polypeptide can be derived from a, wild-type fluorescent polypeptide (i.e., a reference polypeptide) by random or site-directed mutagenesis, including insertions, deletions or truncations or fusions. A fluorescent polypeptide derived from a wild-type polypeptide can have different fluorescence characteristics than the wild-type polypeptide.

As used herein, the term “fluorescence characteristic”, refers to a property of the excitation or emission by a fluorescent polypeptide. Fluorescence characteristics include, for example, the wavelength(s) at which a fluorescent polypeptide is excited or at which it emits (including the breadth and amplitudes of the spectra for each), the extinction coefficient or intensity of the emission, quantum yield or the efficiency of emission, and resistance or susceptibility to photobleaching. Table 2 provides examples of excitation maxima, emission maxima, extinction coefficient and quantum yield for a variety of fluorescent polypeptides.

As used herein, the term “spectrum characteristic of a fluorescent acceptor” refers to the emission spectrum of a given fluorophore that is being used as the fluorescence acceptor in an acceptor/donor pair.

Detailed Description of the Invention

5 In one aspect, the invention relates to dimeric fluorescent proteins that avoid the problems caused by heterodimerization. In this aspect, heterodimerization is avoided by fusing two monomers of the fluorescent polypeptide using a linker peptide. The close spatial relationship of the fused monomers strongly favors the formation of a dimer between the two fused monomers, to the essential exclusion of other monomers sharing a similar dimerization
10 interface. The interaction of the fused monomers via their respective dimerization interfaces is referred to herein as “intramolecular dimerization”. An intramolecular dimer fluorescent protein (IDFP) does not comprise fluorescent monomers that are related to each other as donor and acceptor. That is, the monomers that are linked in an IDFP cannot undergo FRET between them. IDFPs may be co-expressed within the same cell or otherwise mixed with distinct fluorescent
15 proteins comprising the same fluorescent protein dimerization interfaces without encountering the problems caused by heterodimer formation.

 In order to make an IDFP, the nucleic acid encoding a monomer of a fluorescent protein is joined in frame at its 3' end to a sequence encoding a peptide linker, which is itself joined in frame to another copy of the nucleic acid encoding the monomer. This sequence may and often
20 will be additionally linked in frame to a sequence encoding a polypeptide of interest, for example, a polypeptide being investigated for interaction with another protein. Translation of the mRNA encoded by such a nucleic acid construct generates the fluorescent monomers in such close proximity to each other that intramolecular homodimerization of the monomers is very

strongly favored over intermolecular heterodimerization. The resulting polypeptide, therefore, comprises an intramolecular homodimer of the fluorescent protein monomers, fused to a protein of interest.

Fluorescent Proteins Useful According to the Invention

5 Any fluorescent protein that homodimerizes in a cell can be useful in generating an IDFP of the invention. GFPs from *Aequorea victoria*, *Renilla reniformis* and *Renilla mulleri*, among others, are homodimers as they exist in nature. Any of these proteins, and any mutants or engineered versions of these proteins that retain the ability to homodimerize may be used to generate an IDFP of the invention.

10 In order to generate an IDFP according to the invention, the fluorescent protein or the natural protein it was derived from (e.g., *R. reniformis* GFP) must form homodimers when expressed in a monomeric form. It is generally known in the field whether a given protein exists as a homo- or heterodimer in vivo or if it has the capacity to homodimerize. In the event that such knowledge is not available, there are a number of ways in which one of skill in the art may
15 determine whether a particular fluorescent protein homodimerizes. First, biophysical methods such as X-ray crystallography, nuclear magnetic resonance, radiation target analysis or mass spectrometry can be used to determine whether a polypeptide dimerizes.

A biochemical approach is to fractionate samples of purified proteins by size selection gel chromatography under denaturing versus non-denaturing conditions and analyze fractions for the fluorescent protein by fluorescence. If the fluorescent protein migrates at a larger size (approximately twice as large) under non-denaturing conditions relative to denaturing conditions, it is an indication that the protein is a dimer under native conditions. Examples of commonly used matrices include, for example, Sephadex (G10-G200), Bio-Gel (P-2 — P-300) and

Sepharose (2B, 4B, etc.) matrices. One of skill in the art may readily, select a size separation matrix appropriate for such analyses. If performed with purified protein this method can indicate whether or not a polypeptide homodimerizes. If the method is applied to non-purified protein, for example, to protein extracts, the assay only indicates that a dimer forms with some polypeptide, and further analysis is required to determine if the dimer is a homodimer.

Another biochemical method of investigating dimer formation is to generate a truncated or elongated form of the protein and mix it, either by co-expression or by mixing of isolated proteins, with the wild-type protein. If homodimers can form, there will be three distinctly sized bands following native gel electrophoresis: 1) a homodimer of the wild-type; 2) a homodimer of the elongated or truncated form; and 3) an intermediate-migrating diagnostic heterodimer complex of the wild-type and the truncated forms. In the absence of dimerization, only bands (1) and (2) will form.

Additionally, homodimer formation is detected by the method of analytical ultracentrifugation (Baird et al., 2000, Proc Natl Acad Sci U S A., 22:11984-9).

Examples of known fluorescent proteins that can be expressed as intramolecular dimers are as follows. SEQ ID NO: 1 (Figure 1) is the nucleotide sequence encoding wild-type rGFP, and SEQ ID NO: 2 (Figure 2) is the amino acid sequence of wild-type rGFP. A preferred embodiment of the IDFP comprises two copies of the wild-type rGFP polypeptide, linked by a peptide linker sequence. Another embodiment encompasses the same rGFP IDFP additionally fused in frame to a protein of interest. Any protein derived from the rGFP of SEQ ID NO: 2 can be used to generate an IDFP of the invention as long as it retains the ability to homodimerize. In a preferred embodiment, the polynucleotide sequence encoding a fluorescent polypeptide (e.g., rGFP of SEQ ID NO: 2) is a humanized polynucleotide rGFP coding sequence, also referred to

herein as hrGFP. Figure 3 shows a humanized polynucleotide sequence (hrGFP) and the rGFP sequence it encodes (SEQ ID Nos: 3 and 4, respectively).

The amino acid and nucleotide sequences of *A. victoria* GFP are known in the art (Prasher et al., 1992, supra) and vectors encoding a variety of mutant *A. victoria*-derived GFPs are also known and are frequently commercially available. For example, Heim et al. (1995, Nature 373: 663-664) teaches mutations at S65 of *A. victoria* GFP that enhance the fluorescence intensity of the polypeptide. The mutant containing the S65T mutation is particularly important, since its fluorescence is approximately 35 times as intense as wild-type *A. victoria* GFP, and its emission spectrum is shifted to the red, making it more amenable to standard rhodamine optics (excitation and emission maxima at 489 nm and 508 nm, respectively). An S65T mutant encoded by a construct comprising humanized codons is known as EGFP, or “enhanced GFP” (available from CLONTECH; see GenBank Accession No. U43284).

The EGFP mutant is the cornerstone of a series of commercially-available GFP mutants that have differing emission spectra and other useful engineered properties (Cormack et al., 1996, Gene 173: 33-38; Yang et al., 1996, Nucl. Acids Res. 24: 4592-4593; Cramer et al., 1996, Nature Biotechnol. 14: 315-319). Each protein in the series contains mutations in addition to the S65T and humanizing mutations, that alter the emission characteristics of the proteins. For example, the cyan fluorescent protein known as ECFP contains six mutations that shift the emission to cyan light (excitation and emission maxima at 434 nm and 477 nm, respectively; see GenBank Accession No. AB041904 and Sawano et al., 2000, Nucl. Acids Res. 28: e78). The blue fluorescent protein known as EBFP contains four mutations that shift the emission spectrum to blue (excitation and emission maxima at 380 nm and 440 nm, respectively). The yellow fluorescent protein known as EYFP (see Ormo et al., Science 273: 1392-1395, clone GFP-10C)

contains mutations shifting the emission, to yellow-green (excitation and emission maxima at 514 nm and 527 nm, respectively). EGFP, ECFP, EYFP and EBFP are all available from CLONTECH.

5 The S65 site has received considerable scrutiny for its role in determining the fluorescence characteristics of the *A. victoria* GFP molecule. Additional mutants at S65 include, for example, S65A, S65C and S65L, each of which have excitation and emission maxima that differ from wild-type *A. victoria* GFP (see Table 2). The nucleotide sequence encoding an S65A mutant is available as GenBank Accession No. U56996. One skilled in the art can introduce mutations necessary to alter S65 to any desired amino acid. Similarly, the additional point
10 mutations detailed in Table 2 can be generated by one of skill in the art.

Other fluorescent proteins useful according to the invention include, for example, *A. victoria*-derived GFPs that are optimized for expression in plants (GenBank Accession No. U87625 and WO 96/27675), are less thermosensitive (GenBank Accession No. U87973), or are more soluble and emit blue fluorescence (GenBank Accession No. U70497). *A. victoria* GFPs
15 targeted to specific organelles have also been described, such as those targeted to the mitochondria and the nucleus (Rizzuto et al., 1996, Curr. Biol. 6: 183-188). This, listing is by no means exhaustive. There are, for example, a number of fluorescent protein variants, both derived from *A. victoria* and from other sources, that have been reported in or are the subject of U.S. and international patents and patent applications, for example, U.S. Patent Nos. 6,124,128,
20 6,066,476, 6,020,192, 5,804,387, 5,874,304, 5,968,738, 5,625,048, and 5,777,079, and PCT Application Nos. WO 98/21355, WO 98/06737, WO 97/20078, WO 97/42320 and WO 97/11094. Fluorescent protein variants are also described in a number of additional publications in the scientific literature, including, for example, Ehrig et al., 1995, FEBS Lett. 367: 163-166);

Surpin et al., 1987, Photochem. Photobiol. 45 (Suppl): 95S; and Delagrave et al., 1995, BioTechnology 13: 151-154. Any and all of the fluorescent proteins taught in these sources and elsewhere are candidates for the generation of IDFPs of the invention., provided that they homodimerize and the sequences encoding them are known.

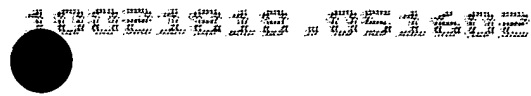
5 The red fluorescent protein from the Indo-Pacific sea anemone of the *Discosoma* species is also a candidate for IDFP generation according to the invention (see Matz et al, 1999, Nature Biotechnol. 17: 969-973). The sequence encoding the protein, known as "DsRed" is available at GenBank Accession No. AF272711, and vectors encoding the protein are commercially available (CLONTECH).

10 Linker Sequences Useful According to the Invention

SW3
B3
15 Linker sequences useful according to the invention serve to join monomers in the dimeric fluorescent polypeptides of the invention. A linker is preferably about 5 to about 50 amino acids in length, and more preferably about 10 to about 20 amino acids in length. An example of linkers useful in the invention are the Gly-Ala linkers taught by Huston et al., U.S. Patent No. 5,258,498, incorporated herein by reference. Additional useful linkers include, but are not limited to (Arg-Ala-Arg-Asp-Pro-Arg-Val-Pro-Val-Ala-Thr)₁₋₅ (Xu et al., 1999, Proc. Natl. Acad. Sci. U.S.A. 96: 151-156), (Gly-Ser)_n (Shao et al., 2000, Bioconjug. Chem. 11: 822-826), (Thr-Ser-Pro)_n (Kroon et al., 2000, Eur. J. Biochem. 267: 6740-6752), (Gly-Gly-Gly)_n (Kluczyk et al., 2000, Peptides 21: 1411-1420), and (Glu-Lys)_n (Kluczyk et al., 2000, supra), wherein n is 20 1 to 15 (each of the preceding references is also incorporated herein by reference).

Proteins of Interest

Frequently it will be advantageous to express an IDFP of the invention as a fusion with a protein of interest. The protein of interest can be any protein for which the nucleic acid sequence



is known and for which that sequence or at least a relevant part of that sequence can be cloned into a vector encoding an IDFP. By relevant part is meant a domain of interest within a protein, for example, a domain being evaluated for protein:protein interactions or a domain with catalytic activity. As used herein, the term “protein of interest” or “domain of interest” refers to any polypeptide or protein, or polypeptide or protein domain, that one wishes to fuse to an IDFP molecule of the invention. The fusion of an IDFP with a polypeptide of interest may be through linkage of the IDFP sequence to either the N or C terminus of the fusion partner. Fusions comprising IDFP polypeptides of the invention need not comprise only a single polypeptide or domain in addition to the IDFP. Rather, any number of domains of interest may be linked in any way as long as the IDFP coding region retains its reading frame and the encoded polypeptide retains fluorescence activity under at least one set of conditions. One non-limiting example of such conditions includes physiological salt concentration (i.e., about 90 mM), pH near neutral and 37°C.

Examples of proteins of interest include, but are not limited to receptors (transmembrane and intracellular) and cell surface proteins, growth factors, signal transduction proteins, transcription factors, structural proteins (e.g., cytoskeletal proteins, nuclear matrix proteins, histones, etc.), extracellular matrix proteins, immunoglobulins, bacterial proteins, plant proteins, viral or phage proteins, enzymes, therapeutic proteins, phosphoproteins, glycoproteins, and lipoproteins.

Production of Intramolecular Dimer Fluorescent Proteins

The production of IDFPs from recombinant vectors may be effected in a number of ways known to those skilled in the art. For example, plasmids, bacteriophage or viral vectors may be introduced to prokaryotic or eukaryotic cells by any of a number of ways known to those skilled

in the art. Examples of useful vectors, cells, methods of introducing vectors to cells and methods of detecting and isolating GFP polypeptides and variants thereof are also described herein below.

1. Vectors Useful According to the Invention.

There is a wide array of vectors known and available in the art that are useful for the expression of IDFPs according to the invention. The selection of a particular vector clearly depends upon the intended use of the polypeptide. For example, the selected vector must be capable of driving expression of the polypeptide in the desired cell type, whether that cell type be prokaryotic or eukaryotic. Many vectors comprise sequences allowing both prokaryotic vector replication and eukaryotic expression of operably linked gene sequences.

10 Vectors useful according to the invention may be autonomously replicating, that is, the vector, for example, a plasmid, exists extrachromosomally and its replication is not necessarily directly linked to the replication of the host cell's genome. Alternatively, the replication of the vector may be linked to the replication of the host's chromosomal DNA, for example, the vector may be integrated into the chromosome of the host cell as achieved by retroviral vectors.

15 Vectors useful according to the invention preferably comprise sequences operably linked to the IDFP coding sequences that permit the transcription and translation of the IDFP sequence. Sequences that permit the transcription Of the linked IDFP sequence include a promoter and optionally also include an enhancer element or elements permitting the strong expression of the linked sequences. The term “transcriptional regulatory sequences” refers to the combination of a
20 promoter and any additional sequences conferring desired expression characteristics (e.g., high level expression, inducible expression, tissue- or cell-type-specific expression, or a combination of these) on an operably linked nucleic acid sequence.

The selected promoter may be any DNA sequence that exhibits transcriptional activity in the selected host cell, and may be derived from a gene normally expressed in the host cell or

from a gene normally expressed in other cells or organisms. Examples of promoters include, but are not limited to the following: A) prokaryotic promoters - E. coli lac, tac, or trp promoters, lambda phage P_R or P_L promoters, bacteriophage T7, T3, Sp6 promoters, B. subtilis alkaline protease promoter, and the B. stearothermophilus maltogenic amylase promoter, etc.; B) eukaryotic promoters - yeast promoters, such as GAL1, GAL4 and other glycolytic gene promoters (see for example, Hitzeman et al., 1980, J. Biol. Chem. 255: 12073-12080; Alber & Kawasaki, 1982, J. Mol. Appl. Gen. 1: 419-434), LEU2 promoter (Martinez-Garcia et al., 1989, Mol Gen Genet. 217: 464-470), alcohol dehydrogenase gene promoters (Young et al., 1982, in Genetic Engineering of Microorganisms for Chemicals, Hollaender et al., eds., Plenum Press, NY), or the TPI1 promoter (U.S. Pat. No. 4,599,311); insect promoters, such as the polyhedrin promoter (U.S. Pat. No. 4,745,051; Vasuvedan et al., 1992, FEBS Lett. 311: 7-11), the P10 promoter (Vlak et al., 1988, J. Gen. Virol. 69: 765-776), the *Autographa californica* polyhedrosis virus basic protein promoter (EP 397485), the baculovirus immediate-early gene promoter gene 1 promoter (U.S. Pat. Nos. 5,155,037 and 5,162,222), the baculovirus 39K delayed-early gene promoter (also U.S. Pat. Nos. 5,155,037 and 5,162,222) and the OpMNPV immediate early promoter 2; mammalian promoters - the SV4O promoter (Subramani et al., 1981, Mol. Cell. Biol, 1: 854-864), metallothionein promoter (MT-I; Palmiter et al., 1983, Science 222: 809-814), adenovirus 2 major late promoter (Yu et al., 1984, Nucl. Acids Res. 12: 9309-21), cytomegalovirus (CMV) or other viral promoter (Tong et al., 1998, Anticancer Res. 18: 719-725), or even the endogenous promoter of a gene of interest in a particular cell type.

A selected promoter may also be linked to sequences rendering it inducible or tissue-specific. For example, the addition of a tissue-specific enhancer element upstream of a selected promoter may render the promoter more active in a given tissue or cell type. Alternatively, or in

addition, inducible expression may be achieved by linking the promoter to any of a number of sequence elements permitting induction by, for example, thermal changes (temperature sensitive), chemical treatment (for example, metal ion- or IPTG-inducible), or the addition of an antibiotic inducing agent (for example, tetracycline).

Regulatable expression is achieved using, for example, expression systems that are drug inducible (e.g., tetracycline, rapamycin or hormone-inducible). Drug-regulatable promoters that are particularly well suited for use in mammalian cells include the tetracycline regulatable promoters, and glucocorticoid steroid-, sex hormone steroid-, ecdysone-, lipopolysaccharide (LPS)- and isopropylthiogalactoside (IPTG)-regulatable promoters. A regulatable expression system for use in mammalian cells should ideally, but not necessarily, involve a transcriptional regulator that binds (or fails to bind) nonmammalian DNA motifs in response to a regulatory agent, and a regulatory sequence that is responsive only to this transcriptional regulator;

Tissue-specific promoters may also be used to advantage with IDFP-encoding constructs. A wide variety of tissue-specific promoters is known. As used herein, the term “tissue-specific” means that a given promoter is transcriptionally active (i.e., directs, the expression of linked sequences sufficient to permit detection of the polypeptide product of the promoter) in less than all cells or tissues of an organism. A tissue specific promoter is preferably active in only one cell type, but may, for example, be active in a particular class or lineage of cell types (e.g., hematopoietic cells). A tissue specific promoter useful according to the invention comprises those sequences necessary and sufficient for the expression of an operably linked nucleic acid sequence in a manner or pattern that is essentially the same as the manner or pattern of expression of the gene linked to that promoter in nature. Any tissue specific transcriptional regulatory sequence known in the art may be used to advantage with a vector encoding an IDFP.

In addition to promoter/enhancer elements, vectors useful according to the invention may further comprise a suitable terminator. Such terminators include, for example, the human growth hormone terminator (Palmiter et al., 1983, supra), or, for yeast or fungal hosts, the TPI1 (Alber & Kawasaki, 1982, supra) or ADH3 terminator (McKnight et al., 1985, EMBO J. 4:2093-2099).

5 Vectors useful according to the invention may also comprise polyadenylation sequences (e.g., the SV40 or Ad5E1b poly(A) sequence), and translational enhancer sequences (e.g., those from Adenovirus VA RNAs). Further, a vector useful according to the invention may encode a signal sequence directing the recombinant polypeptide to a particular cellular compartment or, alternatively, may encode a signal directing secretion of the recombinant polypeptide.

10 A vector useful according to the invention may also comprise a selectable marker allowing identification of a cell that has received a functional copy of the IDFP-encoding gene construct. In its simplest form, the IDFP sequence itself, linked to a chosen promoter may be considered a selectable marker, in that illumination of cells or cell lysates with the proper wavelength of light and measurement of emitted fluorescence at the expected wavelength allows
15 detection of cells that express the IDFP construct. In other forms, the selectable marker may comprise an antibiotic resistance gene, such as the neomycin, bleomycin, zeocin or phleomycin resistance genes, or it may comprise a gene whose product complements a defect in a host cell, such as the gene encoding dihydrofolate reductase (DHFR), or, for example, in yeast, the Leu2 gene. Alternatively, the selectable marker may, in some cases be a luciferase gene or a
20 chromogenic substrate-converting enzyme gene such as the β -galactosidase gene.

IDFP-encoding sequences according to the invention may be expressed either as freestanding polypeptides or as fusions with other polypeptides. It is assumed that one of skill in the art can, given an IDFP nucleic acid sequence, readily construct a gene comprising a sequence

encoding the IDFP fused in frame to one or more polypeptides or polypeptide domains of interest. References teaching methods to do so include Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, and Ausubel et al. (eds.), 1993, Current Protocols in Molecular Biology on CD-ROM, John Wiley & Sons, New York, NY.

A schematic diagram of a vector encoding the transcription unit of one possible embodiment of the invention is shown in Figure 4. In this embodiment, an intramolecular dimer humanized R. reniformis GFP (hrGFP) is encoded on a construct driven by the strong CMV promoter and containing a multi-cloning site (MCS) downstream of the second, or C-terminal copy of hrGFP. A gene of interest is fused at the C-terminus of the hrGFP dimer by insertion in frame into the MCS. A polyadenylation site sequence is included 3' of the MCS to enhance the stability and processing of the transcript generated. The (Gly₄Ser)₂₋₄ linkers shown represent three examples of a linker peptide sequence useful according to the invention and are not meant to be limiting.

a. Plasmid vectors.

Any plasmid vector that allows expression of an IDFP coding sequence of the invention in a selected host cell type is acceptable for use according to the invention. A plasmid vector useful in the invention may have any or all of the above-noted characteristics of vectors useful according to the invention. Plasmid vectors useful according to the invention include, but are not limited to the following examples: Bacterial - pQE70, pQE60, pQE-9 (Qiagen) pBs, phagescript, psiXI74, pBluescript SK, pBsKS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene); pTrc99A, pKK223-3, pKX233-3, pDR540, and pRIT5 (Pharmacia); Eukaryotic - pWLneo,

5

10

2.

20

a suitable host. Vectors suitable for the introduction of IDFP-encoding sequences to host cells from a variety of different organisms, both prokaryotic and eukaryotic, are described herein above or known to those skilled in the art.

Host cells may be prokaryotic, such as any of a number of bacterial strains, or may be eukaryotic, such as yeast or other fungal cells, insect or amphibian cells, or mammalian cells including, for example, rodent, simian or human cells. Host cells may also be plant cells. Cells expressing IDFPs may be primary cultured cells, for example, primary human fibroblasts or keratinocytes, or may be an established cell line, such as NIH3T3, 293T or CHO cells. Further, mammalian cells useful for expression of IDFPs may be phenotypically normal or oncogenically transformed. It is assumed that one skilled in the art can readily establish and maintain a chosen host cell type in culture.

3. Introduction of IDFP-Encoding Vectors to Host Cells.

IDFP-encoding vectors may be introduced to selected host cells by any of a number of suitable methods known to those skilled in the art. For example, IDFP constructs may be introduced to appropriate bacterial cells by infection, in the case of E. coli bacteriophage vector particles such as lambda or M13, or by any of a number of transformation methods for plasmid vectors or for bacteriophage DNA. For example, standard calcium-chloride-mediated bacterial transformation is still commonly used to introduce naked DNA to bacteria (Sambrook et al., 1989, supra), but electroporation may also be used (Ausubel et al. (eds.), supra, 1993).

For the introduction of IDFP-encoding constructs to yeast or other fungal cells, chemical transformation methods are generally used (e.g. as described by Rose et al., 1990, Methods in Yeast Genetics, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY). For transformation of *S. cerevisiae*, for example, the cells are treated with lithium acetate to achieve

transformation efficiencies of approximately 10^4 colony-forming units (transformed cells)/ μ g of DNA. Transformed cells are then isolated on selective media appropriate to the selectable marker used. Alternatively, or in addition, plates or filters lifted from plates may be scanned for IDFP fluorescence to identify transformed clones.

5 For the introduction of IDFP-encoding vectors to mammalian cells, the method used will depend upon the form of the vector. For plasmid vectors, DNA encoding an IDFP may be introduced by any of a number of transfection methods, including, for example, lipid-mediated transfection ("lipofection"), DEAE-dextran-mediated transfection, electroporation or calcium phosphate precipitation. These methods are detailed, for example, in Ausubel et al. (eds.), 1993,
10 supra.

Lipofection reagents and methods suitable for transient transfection of a wide variety of transformed and non-transformed or primary cells are widely available, making lipofection an attractive method of introducing constructs to eukaryotic, and particularly mammalian cells in culture. For example, LipofectAMINETM (Life Technologies) or LipoTaxiTM (Stratagene) kits
15 are available. Other companies offering reagents and methods for lipofection include Bio-Rad Laboratories, CLONTECH, Glen Research, InVitrogen, JBL Scientific, MBI Fermentas, PanVera, Promega, Quantum Biotechnologies, Sigma-Aldrich, and Wako Chemicals USA.

For the introduction of IDFP-encoding vectors to insect cells, liposome-mediated transfection is commonly used, as is baculovirus infection. Cells such as Schneider-2 cells
20 (*Drosophila melanogaster*), Sf-9 and Sf-21 cells (*Spodoptera frugiperda*) or High FiveTM cells (*Trichoplusia ni*) may be transfected using any of a number of commercially available liposome transfection reagents optimized for use with insect cells. Reagents include, for example, TransIT-InsectaTM (PanVera), FuGENETM-6 (Roche), InsectinTM-Plus (InVitrogen) and TfxTM-

20 (Promega). Each of these reagents, used according to the vendor's instructions, permits the introduction of nucleic acid vectors encoding an IDFP to insect cells. Expression vectors optimized for insect cell expression are widely known and are commercially available from, for example, Clontech and InVitrogen. These include both plasmid-based vectors and baculovirus
5 vectors. Insect cell expression vectors are described in detail in "Baculovirus Expression Vectors", D.R. O'Reilly, L.K. Miller & V.A. Luckow (1992, W.H. Freeman Co., New York).

Following transfection with an IDFP-encoding vector of the invention, eukaryotic (preferably, but not necessarily mammalian) cells successfully incorporating the construct (intra or extrachromosomally) may be selected, as noted above, by either treatment of the transfected
10 population with a selection agent, such as an antibiotic whose resistance gene is encoded by the vector, or by direct screening using, for example, FACS of the cell population or fluorescence scanning of adherent cultures. Frequently, both types of screening may be used, wherein a negative selection is used to enrich for cells taking up the construct and FACS or fluorescence scanning is used to further enrich for cells expressing IDFPs or to identify specific clones of
15 cells, respectively. For example, a negative selection with the neomycin analog G418 (Life Technologies, Inc.) may be used to identify cells that have received the vector, and fluorescence scanning may be used to identify those cells or clones of cells that express an IDFP to the greatest extent.

4. Modification of nucleotide sequences to enhance translation of IDFPs.

20 In many applications it will be advantageous to enhance the expression of fluorescent proteins derived from marine invertebrates or bacteria by modifying the codons in the coding sequences to make them more compatible with codon usage in higher eukaryotes, such as mammals and humans. The methods for this so-called "humanizing" are known in the art and, as

noted above, have been applied to *A. victoria* GFP and mutants thereof (U.S. Patent Nos. 6,020,192 and 5,874,304). Humanization is accomplished by site-directed mutagenesis of the less favored codons to more highly favored codons for the same amino acid, as described herein or as known in the art. The preferred codons for human gene expression are listed in Table 1.

The codons in the table are arranged from left to right in descending order of relative use in human genes. In particular, those codons underlined in the table are almost never used in known human genes and, if found in a sequence to be humanized, would therefore represent the most important codons to modify for enhanced expression efficiency in mammalian or human cells. A sequence is considered "humanized" if the codon for one or more amino acids has been changed from the native codon sequence to a codon sequence more favored for translation in human or mammalian cells, preferably without altering the polypeptide coding sequence. Site-directed mutagenesis is well known in the art and is often performed using commercially available kits, such as the EXSITE™ (Catalog No. 200502), QUIKCHANGE™ (Catalog No. 200518) or CHAMELEON® mutagenesis kits (Catalog No. 200509), available from Stratagene.

TABLE 1

PREFERRED DNA CODONS FOR HUMAN USE

| Amino Acids | | | Codons Preferred in Human Genes |
|---------------|-----|---|-----------------------------------|
| Alanine | Ala | A | GCC GCT GCA GCG |
| Cysteine | Cys | C | TGT TGT |
| Aspartic acid | Asp | D | GAC GAT |
| Glutamic acid | Glu | E | GAG GAA |
| Phenylalanine | Phe | F | TTC TTT |
| Glycine | Gly | G | GGC GGG GGA GGT |
| Histidine | His | H | CAC CAT |
| Isoleucine | Ile | I | ATC ATT ATA |
| Lysine | Lys | K | AAG AAA |
| Leucine | Leu | L | CTG TTG CTT <u>CTA</u> <u>TTA</u> |
| Methionine | Met | M | ATG |
| Asparagine | Asn | N | AAC AAT |
| Proline | Pro | P | CCC CCT CCA CCG |
| Glutamine | Gln | Q | CAG CAA |

| | | | | |
|---|------------|-----|---|-------------------------|
| | Arginine | Arg | R | CGC AGG CGG AGA CGA CGT |
| | Serine | Ser | S | AGC TCC TCT AGT TCA TCG |
| | Threonine | Thr | T | ACC ACA ACT ACG |
| | Valine | Val | V | GTG GTC GTT <u>GTA</u> |
| 5 | Tryptophan | Trp | W | TGG |
| | Tyrosine | Tyr | Y | TAC TAT |

The codons at the left represent those most preferred for use in human genes, with human usage decreasing towards the right. Underlined codons are almost never used in human genes.

5. Purification of intramolecular dimer fluorescent proteins.

10 Recombinant fluorescent proteins can be purified from bacteria as follows. Bacteria transformed with a recombinant IDFP-encoding vector of the invention are grown in Luria-Bertani medium containing the appropriate selective antibiotic (e.g., ampicillin at 50 µg/ml). If the vector permits, recombinant polypeptide expression is induced by the addition of the appropriate inducer (e.g., IPTG at 1 mM). Bacteria are harvested by centrifugation and lysed by
15 freeze-thaw of the cell pellet. Debris is removed by centrifugation at 14,000 x g, and the supernatant is loaded onto a Sephadex G-75 (Pharmacia, Piscataway, NJ) column equilibrated with 10 mM phosphate buffered saline, pH 7.0. Fractions containing IDFP are identified by fluorescence emission at the expected wavelength when excited by light in the excitation wavelength.

20 If necessary, IDFPs can be isolated from eukaryotic cells by methods well known to those skilled in the art.. It is also contemplated that IDFPs will include a marker or affinity tag sequence to permit affinity purification. Examples include 6X-His, glutathione S transferase (GST), or epitope tags such as Flag or the Myc tag. These tags are useful for both bacterial and eukaryotic cell expression and purification of IDFPs.

as using an unnatural amino acid, such as a D-amino acid, particularly D-alanine, by functionalizing the amino or carboxylic terminus, e.g. for the amino group, acylation or alkylation, and for the carboxyl group, esterification or amidification, or the like.

Candidate agents will be effective at varying concentrations, depending on the nature of the agent and on the nature of its interaction with the polypeptide or polypeptide fragment of interest. Therefore, candidate agents should be screened at varying concentrations. Generally, concentrations from about 10 mM to about 1 fM are preferred for screening. The association constants of agents that bind polypeptides or fragments thereof will generally be in the range of about 1 mM to about 1 fM, and optimally in the range of about 1 μ M to about 1 pM or less.

10 Uses of Intramolecular Dimer Fluorescent Proteins According to the Invention

IDFPs can be used in any application for which fluorescent proteins are suited. For example, IDFPs can be used as reporter genes to monitor the activity of promoter sequences, to investigate the cellular localization of fusion proteins, to mark cellular proteins for FACS analyses of cell populations, to monitor viral vector infection, to monitor transgene expression in vivo or in culture, and to monitor protein:protein interactions both in vivo and in vitro. It is also expected that IDFPs comprising fluorescent proteins whose spectral characteristics are sensitive to intracellular or extracellular environmental changes (e.g., pH, redox. status, phosphorylation of the fluorescent protein, etc.) will continue to be sensitive to those changes in the context of an IDFP. Because IDFPs do not heterodimerize, they are particularly well suited for multiple-labeling studies involving the co-expression of IDFP-fusion proteins with differing spectral characteristics. Techniques useful for the detection of IDFP fusion proteins include, for example, standard fluorescent microscopy, confocal microscopy, flow cytometry and fluorescence activated cell sorting (FACS).

background. In one embodiment, two different IDFPs that have overlapping emission and excitation spectra (i.e., they are donor and acceptor to each other) are used to generate fusions with two different cellular (or viral) proteins or protein domains being investigated for their ability to interact. A specific interaction of the fusion partners will result in a change in the detected emission spectrum from that of the donor to that of the acceptor when a mixture of the two IDFP fusion proteins is irradiated with light that excites the donor fluorophore. This type of assay is readily adapted to a screening format, in which known interactors are exposed to candidate compounds. Detection of a change from the acceptor's emission profile to the donor's emission profile indicates that a candidate compound has disrupted the interaction between the fusion partners. Either of these assays can be performed in vivo or in vitro. An example of a donor/acceptor fluorescent protein pair is P4-3 and S65C or S65T (Table 2; U.S. Pat. No. 5,981,200). Other examples of donor/acceptor pairs of fluorescent polypeptides include, but are not limited to any of S72A, K79R, Y145F, M153A and T203I (excitation λ 395 nm., emission λ 511 nm) as donor, and any of S65G, T203Y (excitation λ 514 nm, emission λ 527 nm) or T203Y/S65G, V68L or Q69K (excitation λ 515 nm, emission λ 527 nm) as acceptor (See Tsien et al., WO 97/28261). Each of these proteins shares the dimerization interface of *A. victoria*

A pair of fluorescent proteins that are useful according to the invention function as fluorescent donor and fluorescent acceptor, respectively, in the context of fluorescence resonance

energy transfer. The ability of a pair of fluorescent proteins to function as fluorescent donor and fluorescent acceptor, respectively, in the context of fluorescence resonance energy transfer, is determined experimentally and is influenced by a number of factors including donor/acceptor peaks, emission/excitation peaks, peak widths, the efficiency of energy transfer within a fluorescent moiety and peak overlap. Preferably, 1) the donor excitation peak, A, figure 5, will overlap minimally with the acceptor excitation peak, C, figure 5, such that excitation of the donor does not excite the acceptor; 2) the donor excitation peak, A, figure 5, and the donor emission peak, B, figure 5, have sufficient overlap to permit efficient energy transfer; 3) the donor emission peak, B, figure 5, and the acceptor excitation peak, C, figure 5, have sufficient overlap to permit efficient FRET energy transfer; 4) the donor emission peak, B, figure 5 and the acceptor emission peak, D, figure 5 have sufficient overlap to allow for differentiation between FRET and non-FRET energy transfer; and 5) the donor excitation peak, A, figure 5 and the acceptor emission peak, D, figure 5 have sufficient overlap to allow for differentiation between FRET and non-FRET energy transfer.

Generally, an acceptable donor/acceptor pair exhibits $\geq 50\%$ quenching of donor emission at a chromophore distance of $\geq 10 \text{ \AA}$. This is based on the Förster radius, R_0 , which is the distance at which 50% of excited donors are deactivated by FRET (i.e., distance at which energy transfer is 50% efficient). The value of R_0 is dependent on the spectral properties of the donor and acceptor fluorophores, with a general formula: $R_0 = [8.8 \times 10^{23} \cdot \kappa^2 \cdot n^{-4} \cdot QY_D \cdot J(\lambda)]^{1/6} \text{ \AA}$, where κ^2 = dipole orientation factor (0-4; $\kappa^2 = 2/3$ for randomly oriented donors and acceptors); QY_D = the quantum yield of the donor in the absence of the acceptor; n = the refractive index; and $J(\lambda)$ = the spectral overlap integral (the shared area under the overlapping donor emission and acceptor excitation peaks).

The advantage resulting from the forced intramolecular homodimer formation is most apparent when, for example, fluorescent proteins with different emission characteristics, derived from the same parent fluorescent protein, are expressed in a single cell. For example, if two variants of *R. reniformis* GFP have spectral characteristics that permit FRET between the variants, both of these proteins will have the same dimerization interfaces. Without the forced homodimerization occurring in an IDFP, the background level of acceptor fluorescence upon irradiation within the donor's excitation spectrum will be higher than if IDFP versions of the same fluorescent proteins are used.

Even if FRET does not occur between differing fluorescent proteins that share the same dimerization interface, heterodimerization between two fluorescent fusion proteins via that interface can be a problem. For example, such heterodimerization can reduce the sensitivity of sub-cellular localization studies using two labels. Heterodimerization will segregate the labeled proteins into three populations: homodimers of the first fusion protein, homodimers of the second fusion protein, and heterodimers comprising both. Even if one assumes, strictly for the sake of argument, that the heterodimerization will not affect the intracellular localization of the proteins, heterodimer formation will reduce the amount of either homodimer available to segregate to a given location in the cell. This will result in decreased sensitivity in the assay. Therefore, the use of IDFPs in such a situation will improve upon detection sensitivity even if one is not relying upon FRET for detection.

EXAMPLES

Example 1. Monitoring the interaction of polypeptides using IDFPs.

IDFPs are well suited for applications that monitor the association of fusion polypeptides using energy transfer. In order to monitor the association of two polypeptides of interest using

IDFPs, one must first select a pair of fluorescent polypeptides that are donor and acceptor to each other. Each polypeptide in the pair must be capable of homodimerization. An example of such a pair P4-3 and S65T. Another pair useful according to the invention is P4-3 and *R. reniformis* GFP (hrGFP). The nucleic acid sequence encoding the fluorescence donor polypeptide is used to generate a construct encoding, in order, a copy of the donor polypeptide (e.g., P4-3), a linker, a second copy of the fluorescence donor polypeptide and one of the polypeptides of interest (alternatively, the sequence encoding the protein of interest may be placed upstream of and in frame with the sequences encoding the IDFP). Similarly, the sequence encoding the fluorescence acceptor polypeptide is used to generate a construct encoding, in order, a copy of the acceptor polypeptide (e.g., S65T), a linker, a second copy of the fluorescence acceptor polypeptide and the second polypeptide of interest (alternatively, the sequence encoding the protein of interest may be placed upstream of and in frame with the sequences encoding the acceptor IDFP). An example of a pair of proteins of interest is the Ras proto-oncogene product and the Raf-1 kinase. The G-Protein Ras binds to Raf-1 in response to signals originating at receptor tyrosine kinases. A human c-Ha-Ras cDNA sequence is available at GenBank Accession No. J00277 , and a human Raf-1 kinase sequence is available at GenBank Accession No. NM002880. As an example, Ras coding sequences may be ligated in frame to the donor P4-3 IDFP construct, and the Raf-1 coding sequences may be ligated in frame to the acceptor S65A IDFP construct.

Constructs encoding the two IDFP-fusion proteins are transfected, either simultaneously or sequentially into cells in which the protein:protein interaction is to be studied (e.g., HeLa cells, NIH3T3 cells, or another specific cell type of interest) using methods well known in the art

(e.g., lipofection, electroporation, calcium phosphate precipitation, or even retroviral infection following generation of recombinant retroviral vector particles as known in the art).

After selection of cells incorporating and expressing the constructs by standard methods, the interaction of the proteins of interest is measured by detection of fluorescent emission upon
5 irradiation with light that excites the donor fluorophore, in this instance P4-3, but not the acceptor fluorophore, S65T. If the fused Ras and Raf-1 domains interact, excitation with 381 nm light will result in energy transfer between the P4-3 and S65T fluorophores and emission of light with a maximum at about 511 nm. In contrast, if the domains do not interact, the emission maximum upon excitation at 381 nm will be at about 445 nm, the emission maximum of P4-3.
10 This therefore allows the monitoring of the interaction of the two domains in response to stimuli, such as the addition of growth factor, growth factor analogs, or candidate modulators of the signal transduction pathway.

The protein:protein interaction assay using IDFP fusion proteins described above may also be performed in vitro with isolated or purified IDFP fusion proteins. This type of assay, or
15 even the cell-based assay described above may be readily adapted to a high-throughput format by placing the transfected cells or protein samples in a multiwell container and monitoring fluorescence output of samples exposed to various candidate modulators. Further, by performing the interaction assay in the presence or absence of a candidate modulator, one may adapt the method for screening of candidate modulator compounds to identify compounds that either
20 increase or decrease the measured interaction. A change in the interaction in the presence of a candidate modulator relative to the interaction in its absence is indicative of a modulatory effect.

Example 2. Labeling a cell with an IDFP.

IDFPs according to the invention can be used in any application in which fluorescent polypeptides are useful. For example, cells can be labeled by expression of IDFPs to monitor the uptake and expression of transgene constructs, including plasmid-based and retroviral constructs.

5 Cells may also be labeled to facilitate subsequent FACS analysis in a mixed population. To label a cell, an IDFP-encoding construct is introduced to cells by standard methods appropriate to that cell type. Following introduction, selection for cells receiving the construct can either be performed by standard positive or negative selection based on additional selectable marker sequences (e.g., antibiotic resistance genes), by sorting or selection by FACS, or by allowing
10 cells to form colonies and isolating those colonies that fluoresce when irradiated with light within the excitation spectrum of the IDFP. Maintaining the cells under conditions permitting the expression of the IDFP will permit the detection of the cells by fluorescence.

Example 3. Double label monitoring of protein localization.

Fluorescently labeled proteins are often used to examine the sub-cellular localization of
15 proteins of interest. Frequently, it is useful to monitor the localization of two or more proteins or protein domains simultaneously, for example, as a means of identifying relationships between the proteins. When two or more proteins are labeled with fluorescent polypeptides that have the capacity to heterodimerize, the sensitivity of the localization assay can be adversely affected by heterodimerization between the fluorescent polypeptides.

20 Examples of proteins to be monitored for localization include proteins that are recruited to the vicinity of the plasma membrane upon a stimulus such as growth factor engagement of a receptor (e.g., G-proteins, Protein Kinase A, SH2-domain containing proteins, etc.), proteins that

localize to the nucleus in response to a stimulus (e.g., steroid hormone receptor), or proteins that localize to the golgi, mitochondria, nuclear pores or any other subcellular locale.

In order to simultaneously monitor the localization of two proteins of interest, two IDFP fusion constructs, each comprising sequences encoding one of the proteins of interest, are introduced to cells, either simultaneously or sequentially, using standard methods appropriate for that cell type. The localization of the IDFP-tagged proteins is monitored by fluorescence microscopy using excitation wavelengths and filter sets appropriate for the different fluorophores. While not wishing to exclude the possibility, it is generally not necessary that the two IDFPs be fluorescent donor and acceptor to each other. More frequently, unless one is assaying for direct interaction of the proteins, it is preferred that the fluorescent proteins are not related to each other in this manner. The IDFP fusion protein constructs are made using standard methods well known in the art. Examples of pairs of fluorescent polypeptides that are well suited for simultaneous monitoring of localization include, but are not limited to any of S72A, K79R, Y145F, M153A and T203I (excitation λ 395 nm, emission λ 511 nm) as donor, and any of S65G, T203Y (excitation λ 514 nm, emission λ 527 nm) or T203Y/S65G, V68L or Q69K (excitation λ 515 nm, emission λ 527 nm) as acceptor (See Tsien et al., WO 97/2826 1). See also Table 2.

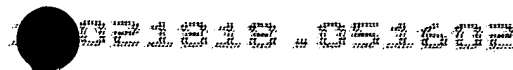
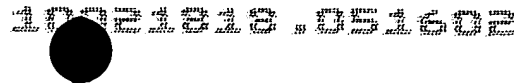


TABLE 2

FLUORESCENCE CHARACTERISTICS OF VARIOUS GFP MUTANTS

| Clone | Mutation(s) | Excitation max (nm) | Emission max (nm) | Extinct. Coefficient (M ⁻¹ cm ⁻¹) | | Quantum Yield |
|-----------|-------------|------------------------|----------------------|--|---------|------------------|
| Wild type | None | 393 (475) | 508 | 21,000 | (7,150) | 0.77 |
| P4 | Y66H | 383 | 447 | 13,500 | | 0.21 |
| P4-3 | Y66H | 381 | 445 | 14,000 | | 0.38 |
| | Y145F | | | | | |
| W-7 | Y66W | 433 (453) | 475 (501) | 18,000 | | 0.67 |
| | N146L | | | (17,100) | | |
| | M153T | | | | | |
| | V163A | | | | | |
| | N212K | | | | | |
| W2 | Y66W | 432 (453) | 480 | 10,000 | (9,600) | 0.72 |
| | I123V | | | | | |
| | Y145H | | | | | |
| | H148R | | | | | |
| | M153T | | | | | |
| | V163A | | | | | |
| | N212K | | | | | |
| S65T | S65T | 489 | 511 | 39,200 | | 0.68 |
| P4-I | S65T | 504 (396) | 514 | 14,500 | (8,600) | 0.53 |
| | M153A | | | | | |
| | K238E | | | | | |
| S65A | S65A | 471 | 504 | | | |
| S65C | S65C | 479 | 507 | | | |
| S65L | S65L | 484 | 510 | | | |
| Y66F | Y66F | 360 | 442 | | | |
| Y66W | Y66W | 458 | 480 | | | |
| 10c | S65G | 513 | 527 | | | |
| | V68L | | | | | |
| | V72A | | | | | |
| | T203Y | | | | | |
| W1B | F64L | 432(453) | 476 (503) | | | |
| | S65T | | | | | |
| | Y66W | | | | | |
| | N146I | | | | | |
| | M153T | | | | | |
| | V163A | | | | | |
| | N212K | | | | | |
| Emerald | S65T | 487 | 508 | | | |
| | S72A | | | | | |



| Clone | Mutation(s) | Excitatin max (nm) | Emission max (nm) | Extinct. Coefficient (M ⁻¹ cm ⁻¹) | | Quantum Yield |
|----------|-------------|-----------------------|----------------------|--|--|------------------|
| | N149K | | | | | |
| | M153T | | | | | |
| | I167T | | | | | |
| Sapphire | S72A | 395 | 511 | | | |
| | Y145F | | | | | |
| | T203I | | | | | |